Nucleic Acids Research, 2010, Vol. 38, Web Server issue ***W175–W181***
doi:10.1093/nar/gkq544

BSQA: integrated text mining using entity relation semantics extracted from biological literature of insects

Xin He[1], Yanen Li[1], Radhika Khetani[2], Barry Sanders[2], Yue Lu[1], Xu Ling[1], ChengXiang Zhai[1] and Bruce Schatz[1,2,*]

[1]Department of Computer Science and [2]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, IL, 61801, USA

Received February 22, 2010; Revised May 20, 2010; Accepted May 29, 2010

ABSTRACT

Text mining is one promising way of extracting information automatically from the vast biological literature. To maximize its potential, the knowledge encoded in the text should be translated to some semantic representation such as entities and relations, which could be analyzed by machines. But large-scale practical systems for this purpose are rare. We present BeeSpace question/answering (BSQA) system that performs integrated text mining for insect biology, covering diverse aspects from molecular interactions of genes to insect behavior. BSQA recognizes a number of entities and relations in Medline documents about the model insect, *Drosophila melanogaster*. For any text query, BSQA exploits entity annotation of retrieved documents to identify important concepts in different categories. By utilizing the extracted relations, BSQA is also able to answer many biologically motivated questions, from simple ones such as, which anatomical part is a gene expressed in, to more complex ones involving multiple types of relations. BSQA is freely available at http://www.beespace.uiuc.edu/QuestionAnswer.

INTRODUCTION

The proliferation of biological literature creates a challenge for individual researchers to keep up with their existing interests, while the paradigm of systems biology encourages researchers to expand their research scope and thinking. These trends significantly increase the information load. Computational processing of a large amount of literature, or text mining as it is often called, promises to relieve these burdens by automatically extracting information from documents (1–3). Information retrieval (IR) methods are developed to retrieve documents or sentences relevant to specific information needs or summarize documents using keywords. These methods have been useful in a number of situations, from aiding database curators to locate papers (4), to interpreting gene lists (5,6). Generally, these methods do not attempt to extract deep semantics from text; instead, they use statistical patterns of words to achieve the goals. In contrast, information extraction (IE) methods specifically aim to identify semantics in the text, often in the form of biological entities and how they are related to each other (relations). IE techniques have been successfully applied to study different relations, from protein–protein interactions (7,8) to gene–disease associations (9).

Both IR and IE methods have limitations. Because IR techniques effectively ignore semantics of terms, it is difficult for them to address questions naturally asked by biologists, even simple ones such as, 'Where is a gene expressed?' While IE methods do attempt to reconstruct meaning from natural language, they are often limited by the need of manually created training data or linguistic rules. As a result, only a small number of entities and relations have been studied, often focused on genes and interactions among genes/proteins, and an even smaller number of systems exist for practical uses.

To make a system practically useful, it is important to cover multiple aspects of the relevant biological domain. For instance, while text mining researchers spent large efforts to optimize the techniques for extracting protein interactions, a biologist may need information about many more aspects such as where the protein is expressed, how it is related to the phenotype of the organism, etc.

*To whom correspondence should be addressed. Tel: (217) 244 0651; Fax: (217) 265 6800; Email: schatz@illinois.edu

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

This trend of integrating multiple types of information to make inference has been recognized well in systems biology research (10,11), but few text mining systems achieve this function.

BeeSpace is the flagship bioinformatics project in the National Science Foundation (NSF) Frontiers of Integrative Biological Research (FIBR) program, see www.beespace.uiuc.edu. The overall goal of BeeSpace is to develop new technologies for functional analysis of genes related to insect behavior, particularly focusing on the honey bee (12). In this work, we present a text mining system for insect biology, as part of BeeSpace. The core component of our BeeSpace question/answering (BSQA) system is the extraction of knowledge in the literature, in the form of various entities, such as genes and anatomical parts, and their inter-relationships. Built on top of this rich representation are two different ways of extracting information. First, for a text query, we automatically identify and rank the entities that appear in the retrieved documents. The ranked list, thus, serves as a compact summary of the documents. As one scenario, a user may query for a biological process, and the returned gene list would suggest genes likely involved in this process. Second, the various relations we recognize from literature are organized in a relational database, and we support a number of queries on this database. Thus a question from a user, such as, 'in what anatomical part is a gene expressed' can be formulated and executed as a structured query language (SQL) query. By utilizing both statistical patterns of entities (our first subsystem) and semantic relations (our second subsystem), we combine the strengths of IR and IE techniques to provide maximum flexibility of information access. Meanwhile, by integrating information on a number of entities and relations, our system enables a user to ask his or her questions from different perspectives.

Comparison to related work

The Textpresso system also annotates various entities, such as genes, in text (4). There are fundamental differences between Textpresso and BSQA. Textpresso is primarily an enhanced IR system, where the queries are fixed sentence templates and the results are sentences and documents to be read by users. In contrast, BSQA performs relation extraction and supports many types of queries modeled on realistic biological questions, as explained above. The results of BSQA are entities and relation instances, which are easier to understand than long lists of documents, saving valuable user effort by automatically extracting the facts within the sentences. There are only a few systems that do practical IE on multiple types of relations, including for instance, PLAN2L for plant biology (13) and STITCH for protein–chemical interactions (14). Beyond the difference in the intended biological domains, these systems do not offer extensive queries. In the domain of insect biology, FlyMine integrates different types of genomic data and supports many relational queries, similar to ours (15). However, FlyMine must rely on experimental data or facts manually extracted from literature by database curators, whereas we automatically extract the relations from literature using text mining techniques, by a process similar to a curator assistant.

METHODS

The flowchart of the BSQA system is shown in Figure 1. The system has two types of modules: those that provide textual data and annotations (the central column of Figure 1), and those that answer user queries (the right column of Figure 1). At the first step, we used a collection of 38 844 abstracts from Medline and Biosis, which were given to us by the FlyBase curators in 2007 as constituting the official collection from which they had extracted facts for gene annotation (see the BeeSpace production software on website for the information of the most recent collection—we are conducting regular updates of the collections). The abstracts are indexed and tokenized by a customized program using Lemur toolkit, which normalizes some special symbols and preserves the integrity of biological entities (16). For example, a hyphen symbol will be removed if it appears between a word and a digit (e.g. brca-1 will be converted to brca1). At the next step, four types of entities are recognized in the documents and marked up in the XML format: *Gene, Anatomy* (tissues or body parts), *Chemicals* and *Behavior*. Genes are recognized by matching words or phrases in documents with official gene symbols as well as their synonyms in FlyBase (case-insensitive string matching).

Since many fly gene names may be ambiguous, e.g. for (foraging), in (inturned), similar (sima), we developed a machine learning method to disambiguate each mention of a gene name according to its context. The ambiguity of a gene name is defined according to whether it appears in a dictionary of English words and common biological terms. The goal of this method is to classify ambiguous gene mentions as gene sense (positive) or non-gene sense (negative). We observe that the majority of gene names in Fly are unambiguous; and the majority of ambiguous gene mentions in the text are negative. We assume that the positive examples in ambiguous gene mentions follow the same feature distribution as the unambiguous gene mentions. We thus treat all unambiguous gene name mentions of Fly as positive examples, and all ambiguous gene mentions as negative examples. This allowed us to train a Naïve Bayes classifier on the contexts of each gene mention, using features such as word distribution in the neighboring window and part-of-speech tagging of the word. The details of this procedure can be found in our website. Our gene name recognition procedure achieves precision at 0.76 and recall at 0.62 in our manual evaluation of 99 randomly chosen abstracts. The entity Anatomy is recognized using the controlled vocabulary of anatomical structure from FlyBase. This simple scheme leads to a high precision (0.98) and recall (0.91) in our evaluation of 103 randomly sampled abstracts. We manually curated a list of chemicals that may affect animal behavior, including neurotransmitters, hormones and secondary messengers. Since no standard naming convention

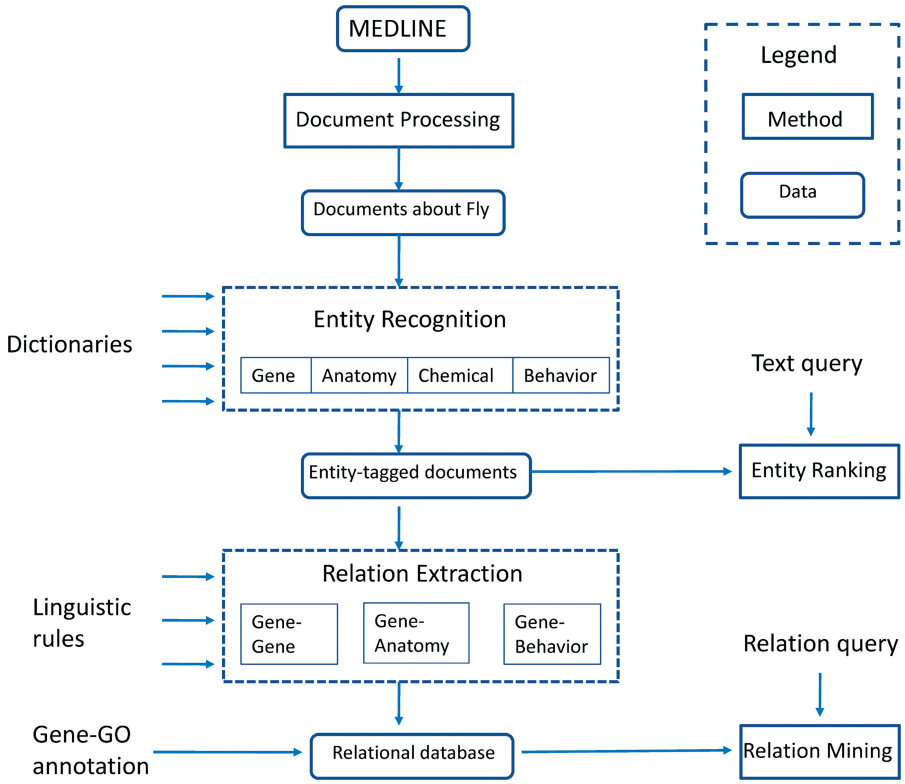

Figure 1. The flowchart of the BSQA system. The main steps of the computational procedure are shown (see text for details).

exists for describing behavior, our strategy is to recognize all bigrams ended with the word 'behavior', and two biologist experts manually chose the behavior terms from this list (e.g. 'foraging behavior' is chosen, but not 'complex behavior'). This strategy may miss a number of terms, but the final list of 748 terms still covers a large range of behavior.

Our next main step is to extract three types of relations from text (Figure 1): Gene–Gene (the first gene regulates the expression of the second gene), Gene–Anatomy (the gene is expressed in the anatomical part or tissue) and Gene–Behavior (the gene plays functional role in the behavior). Because of the lack of training data, our extraction is based on hand-crafted patterns or keywords. Specifically, to extract the Gene–Gene relation, we created a set of regular patterns. For instance, a simple pattern, 'expression of B [GAP] regulated by [GAP] A', will lead to identification of A as the regulator and B as the target, where A and B are recognized gene names, and [GAP] represents a gap of a specified length. Our patterns cover the cases where the relation is explicitly mentioned (the example above), as well as the other cases where the relation can only be inferred (e.g. the promoter of one gene contains a binding site of another gene). The carefully constructed list of 32 patterns (available in the website) achieves precision at 0.65 in our evaluation of a sample data set (64 out of 99 predicted Gene–Gene relations are correct). We followed a procedure similar to that used by Saric and Bork (17) to evaluate recall. This gives us recall at 0.24, slightly below that of Ref. (17) at 0.30. Note that some misses are due to the problems of gene name recognizer (excluding this effect would lead to a recall of 0.33). Considering the fact that gene name recognition is significantly harder in fruit fly than in yeast, the model organism used in Ref. (17), we think the results in the two studies are comparable. The Gene–Anatomy relation is recognized by the keywords appearing in the sentences where a gene name and an anatomical part co-occur. The keyword list includes words such as expression and localization (the full list of 31 keywords is available in the website). Even though the method is simple, we find that in 58 out of 85 predicted relation instances (precision 0.68), the expression relations identified are correct. In a randomly sampled set of 100 abstracts, the program recovered 23 out of 55 total Gene–Anatomy instances, giving recall at 0.42. For Gene–Behavior relation, we reasoned that in most cases where a gene and a behavior term co-occur within a single sentence, there should be some functional relationship between the two, so our extraction is based on co-occurrence. The precision of this procedure is 0.55 (55 out of 100 predicted Gene–Behavior instances, randomly selected, are correct). We did not evaluate recall in this case, since recall should be 100% by definition, if exclude the errors of entity recognition (for any true Gene–Behavior instance, the two entities should co-occur in the same sentence). To enhance our power of answering questions, we also imported the gene ontology (GO) annotation of genes from FlyBase, as Gene–GO relation, into our system. We built a SQL database to store all instances of these four types of relations, as well as other necessary data, e.g. the bibliographical information of articles.

We presented the details of our evaluation procedure and results (for all entities and relations discussed above) in the BSQA website, along with the data we manually created. We built two applications on top of the infrastructure just described. The Entity Ranking component (Figure 1) first retrieves documents relevant to a text query using the built-in capability of Lemur, and then ranks entities according to the frequency of an entity in the relevant documents. The Relation Mining component (Figure 1) maps a user's question, from a predefined list of template questions, to a SQL query, and executes the query on the SQL database.

The system runs on a desktop tower server, in the Institute for Genomic Biology, equipped with 4 quad-core Intel processors (Q6600, 2.4 GHz) and 8 GB RAM. For the backend software infrastructure, the applications are hosted in Apache web server, and the MySql (version 5) is used to power the database service; in the frontend, an AJAX JavaScript framework called EXT JS (version 3) is deployed for the web interface. BSQA is freely available at http://www.beespace.uiuc.edu/QuestionAnswer.

SOFTWARE USAGE

We first describe the usage of our Entity Ranking subsystem. A user types in his free-text query in the search box, and the retrieved documents will be displayed in the main screen, sorted by relevance (Figure 2A). In the results, the entities are highlighted with different colors (the color code is shown alongside the search box), and for each entity identified, a hyperlink is created pointing to an external page explaining the entity or providing more information (e.g. FlyBase gene entry). To gain a quick picture of what concepts may be important in the retrieved documents, a user could inspect the top concepts in each entity category. The entities in the results are sorted by their frequencies in the retrieved documents, and the PMIDs of the supporting documents will be shown to facilitate further investigation (Figure 2B).

To use our Relation Mining subsystem, a user first needs to choose a query template from a predefined list in the pull-down menu, and then type in the variable(s) specific for a query in the corresponding box(es). These templates are designed to model the questions commonly asked by a biologist. Some templates of simple queries are:

- Find the anatomical parts where a gene X is expressed.
- Find the target genes regulated by a gene X.
- Find all genes that may be related to the behavior X.

The symbol X represents a query variable to be input in the query box. In addition, we support some complex queries that may require joining multiple relations. For example:

- Find genes expressed in the anatomical part X and annotated with GO term Y.
- Find pairs of regulator-target genes that are expressed in anatomical tissue X.

The full list of all supported queries can be found in our website. In all cases, the results of a query are a list of entities being searched for, and for each entity in the result, its supporting documents will be displayed, with all the recognized entities in the documents highlighted and hyperlinked as before (Figure 3).

CASE STUDIES

We tested the two functions of BSQA, and present several examples here.

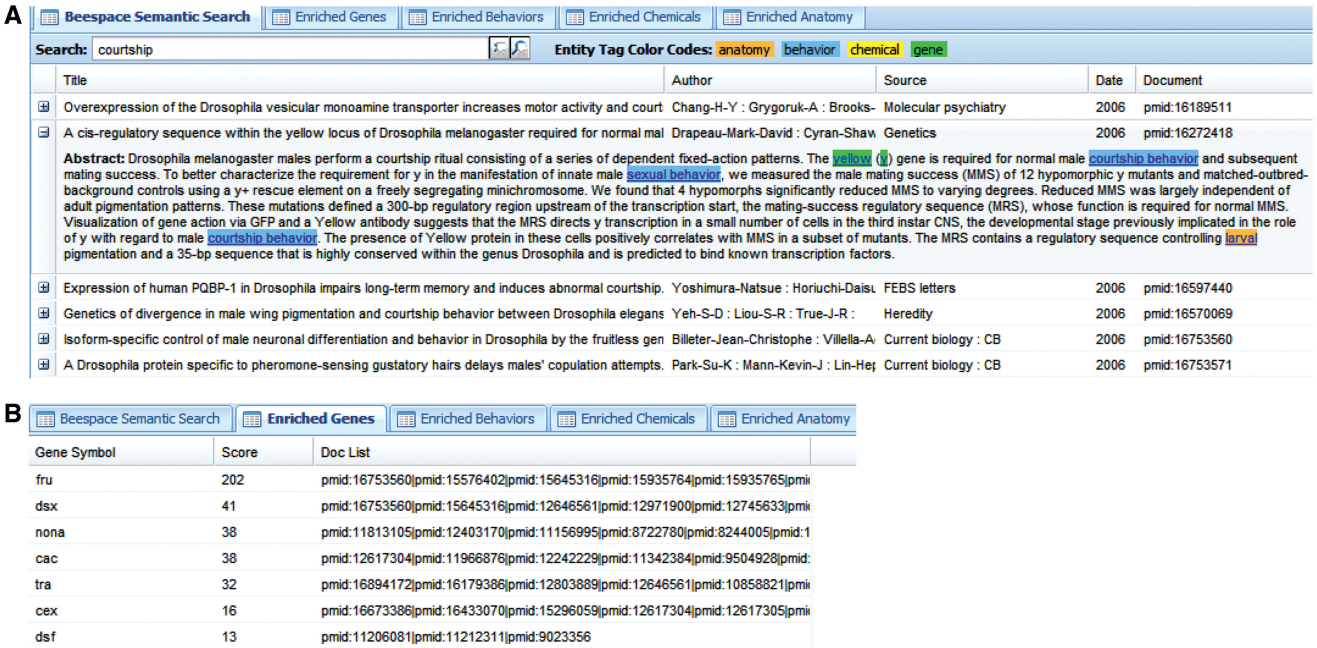

Figure 2. The Entity Ranking subsystem of BSQA. (**A**) The retrieved documents of the example query 'courtship'. Clicking on the title of one result entry will expand its abstract, highlighting entities with different colors. The hyperlinks in the entities point to external resources. (**B**) The genes appearing in the retrieved documents, ranked by their frequencies.

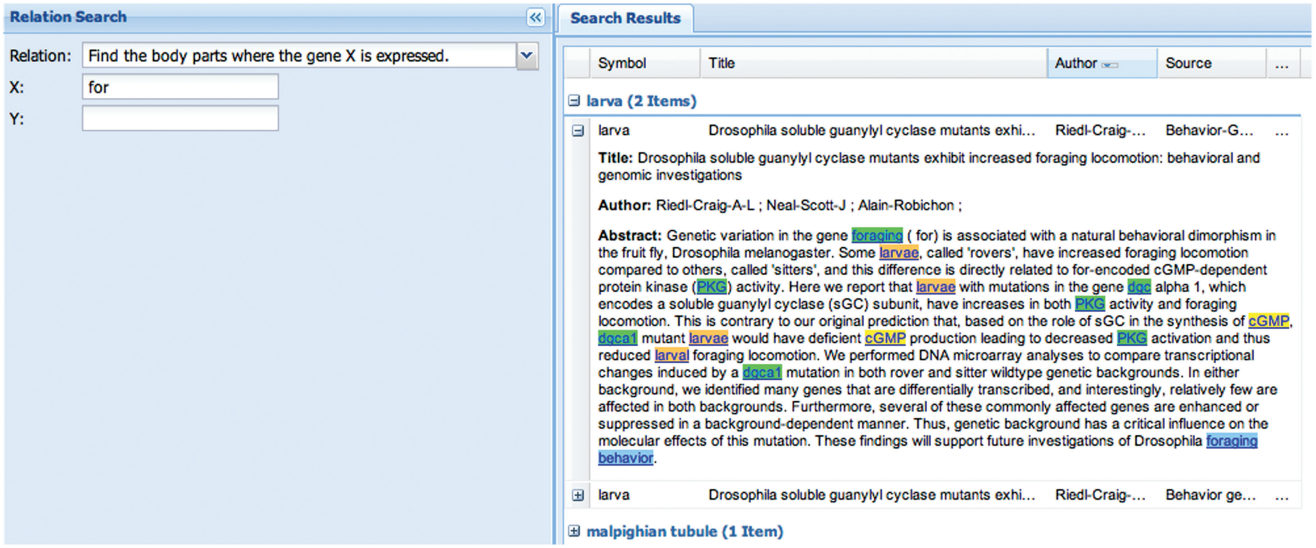

Figure 3. The Relation Mining subsystem of BSQA. In the left panel, a user chooses the template question and types in the query variable(s). The main results in the right panel are sorted by the entities and clicking on the entity names will reveal the associated documents.

Entity Ranking

In our first example, we tested if BSQA is able to recognize automatically important concepts related to an arbitrary text query. We reasoned that this feature would be very useful for a researcher who starts to work on an unfamiliar topic. We utilized it to learn more about the 'synaptonemal complex' (SC), a protein structure in eukaryotes. The query generated a list of 25 enriched genes and 13 enriched anatomical parts. Based on the top five enriched anatomical parts: oocyte, chiasma, nurse cell, gonad and spermatocyte, it appears very likely that this structure is present during oogenesis and/or spermatogenesis. Upon further analysis of the abstracts returned in the search, we determined that the SC is found in cells undergoing meiosis (specialized cell division during oogenesis and spermatogenesis), and it is also necessary for chromosomal recombination taking place during meiosis. The enriched gene list and the supporting documents were used for further in-depth analysis. We confirmed that of the 25 genes in the list of enriched genes, the top 11 genes and a total of 19 genes were involved in the normal structure-function of the SC in Drosophila (18). These results show that the Entity Ranking function of BSQA is effective in suggesting concepts in different categories related to a query, and that these concepts reflect biological findings in literature.

Relation Mining

We next examined the Relation Mining function of BSQA. We started with the query, 'Find all body parts where the gene X is expressed', and tested the gene *bicoid* (*bcd*). The resulting seven anatomical parts summarize the role of *bcd* during Drosophila development (19). The terms such as 'oocyte' and 'ovary' suggest that *bcd* is a maternal gene that is present during oogenesis and the terms 'embryo' and 'pole cell' suggest that *bcd* plays a role in embryogenesis. Further examining the retrieved documents for the term 'oocyte' quickly reveals that *bcd* is localized at the anterior pole of the oocyte. And examining the documents for the term 'embryo' suggests that the maternally deposited *bcd* directs the establishment of anterior–posterior axis in early development. Thus, by inspecting the query results and the supporting documents, one can easily obtain a molecular picture of the expression pattern of a query gene, *bcd* in this example.

Next, we tested the query, 'Find all genes that may be related to the behavior X', with X being 'foraging behavior'. The systems returns five genes: *akh*, *csr*, *for*, *loco* and *svr*. Inspection of the associated documents also returned quickly confirmed that *csr* and *for* influence larval foraging behavior (20,21), and *akh*, as a neuropeptide, influence starvation-induced foraging behavior by regulating the metabolism of the fly (22). Notably, our system correctly identifies the ambiguous gene name *for* in the text, while ignoring the word 'for' as prepositions. The gene *loco* is a false positive because the term 'locomotion defects', a synonym of *loco*, appears in the text discussing foraging behavior, and similarly *svr* is a false positive because its synonym 'cc' is also an abbreviation of 'central complex' (our gene recognizer failed in this case because in this context, 'cc mutants', cc does look like a gene name). This example demonstrates the utility of BSQA to quickly extract information about the genetic basis of a complex behavior, and also illustrates the power as well as limitations of our gene recognizer.

For our last case study, we tested the complex query 'Find the genes that are expressed in body part X and annotated by the GO term Y'. We were interested in finding genes involved in muscle development that are specifically present early in development, in the larval imaginal discs. So we set X to 'imaginal disc' and Y to 'muscle organ development' (GO: 0007517). The system returns three genes: *ap*, *dr* and *ewg*. Closer inspection of the function of these genes on FlyBase reveals that *ap*

(*apterous*) and *ewg* (*erect wing*) are involved in muscle organ development during the larval stage, as expected. The third gene abbreviation *dr* is actually for the gene *Dr* or *Drop* (gene symbols are case-sensitive, while BSQA text processing removes cases), which has also been implicated in muscle development during the larval stage. This example clearly demonstrates the ability of BSQA to answer complex questions that require integrating information from multiple sources.

DISCUSSION

Given the large size of biological literature, how to quickly locate information related to specific questions is a long-term challenge facing biological researchers. In this work, we built a text mining system that aims to address this challenge for insect biologists. Our system extracts various entities and relations automatically from text that capture important aspects of insects at both molecular and organism level. Together these representations allow a researcher to access information relevant to a problem from different viewpoints, and integrate information distributed in different sources. Our system provides maximum flexibility of information access through the use of different query interfaces and a number of biologically motivated query templates. We demonstrated the utility of this system through realistic examples.

One major advantage of BSQA is its expandability. New query templates can be easily added to the existing list of the Relation Mining subsystem. Future user feedbacks will be an important source of new queries. Furthermore, our relational database can easily import relations from other, perhaps, non-text sources, e.g. protein interaction data from high-throughput experiments. We illustrated this feature with GO annotation in this work. The new relations can be joined with the existing ones to support queries using both literature and genomic data.

The current system uses the fruit fly literature as the underlying data source. Because the fly is the model organism for all insects, our system will be useful for most insect biologists. To extend to other insects such as beetles or wasps would be straightforward, as the basic entities (Genes, Anatomy and Behavior) are highly conserved across all insects. We have already produced good preliminary results with a comprehensive insect text collection comprised of 100K Biosis abstracts, while collaborating with the Arthropod Base Consortium for insect genomes and beyond. Another interesting direction is to develop a system with similar functions for other organisms, such as supporting mammals by using mouse as the model organism. This would leverage a different dictionary (MGI) with quality entities, while using similar training sets. Many of our ideas, such as the flexible querying systems, and much of the infrastructure, from relational database to the Web interface, can be applied to new domains. Because of the generality of the design of our system, we could add more entities and relations to deal with a different or more complex biology of different organism. Thus we expect such extensions to other organisms and other functions to be straightforward.

ACKNOWLEDGEMENTS

We would like to thank Moushumi Sen Sarma, Gene Robinson and other BeeSpace members for many helpful discussions during development and testing. The main BeeSpace programmer David Arcoleo helped enhance the software, especially the user interface. FlyBase kindly provided the test source collection and many helpful interactions with their curators, through William Gelbart at Harvard University. The Arthropod Base Consortium, through its annual symposium organized by Susan Brown of Kansas State University, provided a forum for displaying and improving the software. Bioinformatics software is available at www.beespace.uiuc.edu, including the production system that supports the Gene Summarizer (12) and the Genelist Analzyer (16), for functional analysis of genome data using special collections of biological literature.

FUNDING

Funding for open access charge: Frontiers of Integrative Biological Research program (grant 0425852) entitled BeeSpace: An Interactive Environment for Analyzing the Nature-Nurture in Societal Roles.

Conflict of interest statement. None declared.

REFERENCES

1. Rzhetsky,A., Seringhaus,M. and Gerstein,M. (2008) Seeking a new biology through text mining. *Cell*, **134**, 9–13.
2. Ananiadou,S., Kell,D.B. and Tsujii,J. (2006) Text mining and its potential applications in systems biology. *Trends Biotechnol.*, **24**, 571–579.
3. Hoffmann,R. and Valencia,A. (2004) A gene network for navigating the literature. *Nat. Genet.*, **36**, 664.
4. Muller,H.M., Kenny,E.E. and Sternberg,P.W. (2004) Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS Biol.*, **2**, e309.
5. Chaussabel,D. and Sher,A. (2002) Mining microarray expression data by literature profiling. *Genome Biol.*, **3**, RESEARCH0055.
6. Jelier,R., Schuemie,M.J., Veldhoven,A., Dorssers,L.C., Jenster,G. and Kors,J.A. (2008) Anni 2.0: a multipurpose text-mining tool for the life sciences. *Genome Biol.*, **9**, R96.
7. Ono,T., Hishigaki,H., Tanigami,A. and Takagi,T. (2001) Automated extraction of information on protein–protein interactions from the biological literature. *Bioinformatics*, **17**, 155–161.
8. Santos,C., Eggle,D. and States,D.J. (2005) Wnt pathway curation using automated natural language processing: combining statistical methods with partial and full parse for knowledge extraction. *Bioinformatics*, **21**, 1653–1658.
9. Korbel,J.O., Doerks,T., Jensen,L.J., Perez-Iratxeta,C., Kaczanowski,S., Hooper,S.D., Andrade,M.A. and Bork,P. (2005) Systematic association of genes to phenotypes by genome and literature mining. *PLoS Biol.*, **3**, e134.
10. Zhu,J., Zhang,B., Smith,E.N., Drees,B., Brem,R.B., Kruglyak,L., Bumgarner,R.E. and Schadt,E.E. (2008) Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat. Genet.*, **40**, 854–861.

11. Yeger-Lotem,E., Riva,L., Su,L.J., Gitler,A.D., Cashikar,A.G., King,O.D., Auluck,P.K., Geddie,M.L., Valastyan,J.S., Karger,D.R. *et al.* (2009) Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. *Nat. Genet.*, **41**, 316–323.
12. Ling,X., Jiang,J., He,X., Mei,Q., Zhai,C. and Schatz,B. (2006) Automatically generating gene summaries from biomedical literature. *Pac. Symp. Biocomput.*, 40–51.
13. Krallinger,M., Rodriguez-Penagos,C., Tendulkar,A. and Valencia,A. (2009) PLAN2L: a web tool for integrated text mining and literature-based bioentity relation extraction. *Nucleic Acids Res.*, **37**, W160–W165.
14. Kuhn,M., von Mering,C., Campillos,M., Jensen,L.J. and Bork,P. (2008) STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Res.*, **36**, D684–D688.
15. Lyne,R., Smith,R., Rutherford,K., Wakeling,M., Varley,A., Guillier,F., Janssens,H., Ji,W., McLaren,P., North,P. *et al.* (2007) FlyMine: an integrated database for Drosophila and Anopheles genomics. *Genome Biol.*, **8**, R129.
16. He,X., Sen Sarma,M., Ling,X., Chee,B., Zhai,C. and Schatz,B.R. (2010) Identifying overrepresented concepts in gene lists from literature: a statistical approach based on Poisson mixture model. *BMC Bioinformatics*, **11**, 272.
17. Saric,J., Jensen,L.J., Ouzounova,R., Rojas,I. and Bork,P. (2006) Extraction of regulatory gene/protein networks from Medline. *Bioinformatics*, **22**, 645–650.
18. Page,S.L. and Hawley,R.S. (2004) The genetics and molecular biology of the synaptonemal complex. *Annu. Rev. Cell Dev. Biol.*, **20**, 525–558.
19. McGregor,A.P. (2005) How to get ahead: the origin, evolution and function of bicoid. *Bioessays*, **27**, 904–913.
20. Pereira,H.S., MacDonald,D.E., Hilliker,A.J. and Sokolowski,M.B. (1995) Chaser (Csr), a new gene affecting larval foraging behavior in Drosophila melanogaster. *Genetics*, **141**, 263–270.
21. Varnam,C.J., Strauss,R., Belle,J.S. and Sokolowski,M.B. (1996) Larval behavior of Drosophila central complex mutants: interactions between no bridge, foraging, and Chaser. *J. Neurogenet.*, **11**, 99–115.
22. Lee,G. and Park,J.H. (2004) Hemolymph sugar homeostasis and starvation-induced hyperactivity affected by genetic manipulations of the adipokinetic hormone-encoding gene in Drosophila melanogaster. *Genetics*, **167**, 311–323.